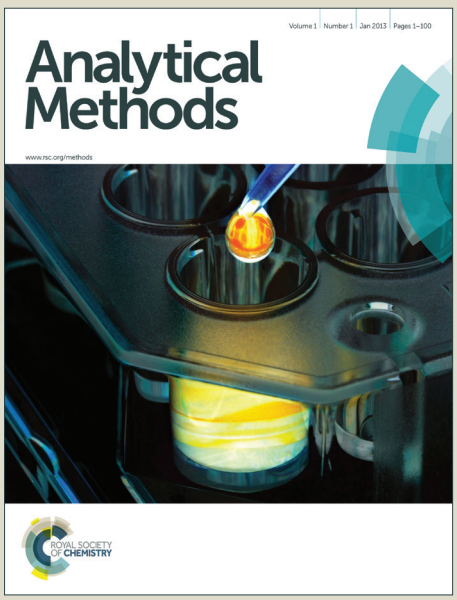
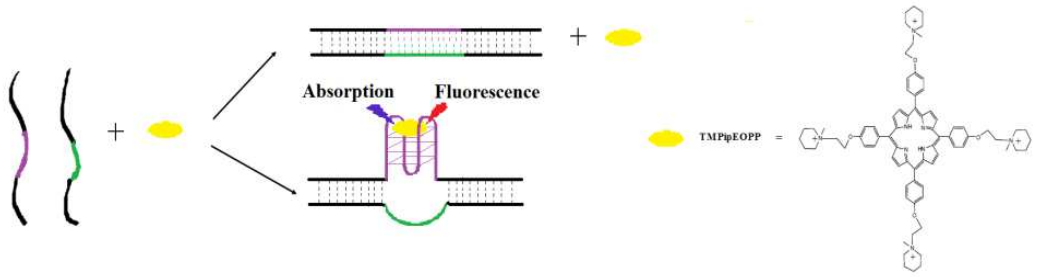
Analytical Methods

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods



A label-free optical method was developed to probe G-quadruplex/duplex competition embedded in flanking duplex DNAs by using a G-quadruplex-specific optical probe TMPipEOPP. Using this method, the formation of G- quadruplexes embedded in flanking duplex DNA can be judged easily and quickly. This method could also be used to monitor conversion between G-quadruplexes and duplexes in real time. Utilizing this method, some important information on G-quadruplex/duplex competition inside duplex DNAs was obtained.

Journal Name

RSCPublishing

ARTICLE

A simple, label-free optical method for studies on G-quadruplex/duplex competition inside duplex DNAs using G-quadruplex-specific probe—TMPipEOPP

Li-Na Zhu,* Yan-Fang Huo and Bin Wu ,

Cite this: DOI: 10.1039/xoxxoooooox

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/xoxxoooooox

www.rsc.org/

G-rich sequences that might form G-quadruplexes are common in some regions of the human genome. Most of these G-rich sequences are embedded in flanking duplex-forming sequences and coexist with complimentary C-rich sequences. Therefore, competition between G-quadruplexes and the duplex structures in flanking duplex DNAs must be considered in corresponding G-quadruplex studies. Based on the high specificity of a porphyrin derivative (TMPipEOPP) for G-quadruplexes over duplexes and single-stranded DNAs, a simple method was developed for studying the competition between G-quadruplexes and duplexes in flanking duplex structures. In the presence of a complementary sequence, core G-rich region of a long DNA sequence tended to form duplex under dilute conditions and form G-quadruplex under molecular crowding conditions. The method could also be used to monitor conversion between G-quadruplexes and duplexes in real time. Under dilute conditions, G-rich sequences rapidly hybridized with complementary strands to form stable duplexes that did not disassociate with time. If the G-rich sequences have been folded into G-quadruplexes, addition of a complementary sequence promoted slow conversion from G-quadruplexes to duplexes. However, under molecular crowding conditions, stable G-quadruplexes formed regardless of the presence of complementary sequences and the formed G-quadruplexes did not disassociate with time. Changing solution conditions from dilute to molecular crowding promoted rapid structural conversion of G-rich sequences from duplexes to G-quadruplexes. This method could be an important tool for G-quadruplex studies.

Introduction

Double-stranded duplex DNA is the most common secondary structure in the human genome. However, in some genomic regions, noncanonical DNA structures can form.[1] A well-known structure is the G-quadruplex, an uncommon DNA secondary structure formed by G-rich DNA or RNA.[2-5] Genomic sequences that have the potential to form G-quadruplexes are widely distributed in the human genome.[6] Research on these genomic regions is important for determining the physiological function of these regions and designing drugs that target G-quadruplex structures.[7-10]

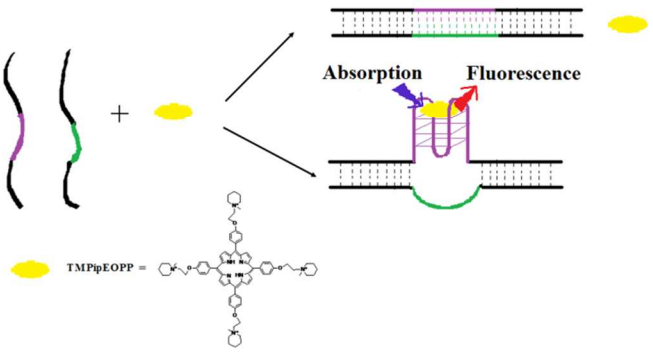
Except for the ~200-nucleotide single-stranded telomeric overhangs at the ends of chromosomes, most G-rich sequences with G-quadruplex-forming potential coexist with their complementary C-rich sequences.[11] Elucidating the physiological functions of these genomic sequences requires understanding the DNA structures formed by the G-rich sequences, specifically whether they fold into G-quadruplex structures or hybridize with their complementary sequences to form duplex helix structures. To resolve these structural questions, competition between G-quadruplex and duplex structures must be investigated.

To ensure that the experimental results can truly reflect DNA structures formed within cells, two factors must be considered. One is the G-rich sequence length; the other is solution conditions. In some studies,[12-16] short, core G-rich sequences were truncated from human DNA and G-quadruplex formation was investigated in the presence of short, complementary C-rich sequences. This study design is inconsistent with the cellular conditions of genomic DNA. In human DNA, the ends of G-rich sequences are usually extended into long sequences that hybridize with complementary sequences to form duplexes. That is, G-rich sequences are embedded in flanking duplex structures. The presence of these flanking duplexes promotes close proximity between G-rich sequences and complementary C-rich sequences, reinforcing the formation of duplexes. Thus, using long DNA sequences containing core G-rich sequences and flanking duplex-forming sequences might yield results from G-quadruplex/duplex competition studies that are more consistent with structures found in cells.[11]

Solution conditions, specifically whether assays are done in dilute or molecular crowding conditions, also affect competition between G-quadruplex and duplex formation. It has been reported that dilute conditions are benefit for the formation of duplexes and molecular crowding conditions destabilize duplexes and stabilize G-quadruplexes.[16-20] Cells are crowded environments in which ~40% of the cellular volume is filled with macromolecules.[20] To obtain results that can reflect cellular G-quadruplex/duplex competition, experiments should be conducted under conditions of molecular crowding.

Using dimethyl sulfate footprinting and gel electrophoresis techniques, competition between G-quadruplexes and duplexes embedded in long duplex DNA was studied by Tan et al..[11] The results demonstrated that stable G-quadruplexes can form in long duplex DNA. Although this method can analyse G-quadruplex formation in long duplex DNA, the interconversion between G-quadruplex and duplex structures cannot be monitored continuously, and relatively long electrophoresis time increases the complexity of the experimental procedure and might allow changes in DNA structure during electrophoresis due to the interaction of DNAs with matrix.[21] Thus, a simple method for investigating G-quadruplex/duplex competition in long DNA molecules is still required.

Optical analysis is promising for investigating DNA structure because it is simple, cost-efficient, easy, and has high sensitivity. G-quadruplex molecular beacon technology has been used in G-quadruplex/duplex competition studies.[16,22,23] The molecular beacon method requires a G-rich oligonucleotide that is end-labeled with a fluorophore on one end and a quencher moiety on the other. The method can be used only to study short G-rich sequences and is not suitable for long DNA sequences in which the quadruplex-forming core G-rich sequence is embedded in flanking duplex-forming sequences. In this study, a simple, label-free method was developed to study G-quadruplex/duplex competition inside long duplex DNA (Scheme 1). This method used a G-quadruplex-specific optical probe (5,10,15,20-tetra-{4-[2-(1-methyl-1-piperidinyl) ethoxy] phenyl} porphyrin, TMPipEOPP). Using changes in the UV-vis absorption or fluorescence spectrum of TMPipEOPP caused by different DNA structures, the formation of G-quadruplex structures in long duplex DNA molecules could easily be determined.



Scheme 1. TMPipEOPP is used as an optical probe to investigate G-quadruplex/duplex competition embedded in duplex DNA

Experimental

Materials and reagents

The oligonucleotides listed in Table 1 were purchased from Sangon Biotech. Co. Ltd. (Shanghai, China). The concentrations of the oligonucleotides were represented as single-stranded concentrations. Single-stranded concentration was determined by measuring the absorbance at 260 nm. Molar extinction coefficient was calculated using a nearest neighbour approximation (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer), and the calculated molar extinction coefficients of these oligonucleotides were listed in Table 1. 5,10,15,20-tetra-{4-[2-(1-methyl-1-piperidinyl)ethoxy] phenyl} porphyrin (TMPipEOPP) was synthesized according to the method reported by our group.[24] Na_2EDTA (disodium ethylenediamine tetraacetic acid), Tris (tris(hydroxymethyl)aminomethane), KCl, $MgCl_2$, polyethylene glycol 200 (PEG 200) were obtained from Sigma. Deionized and sterilized water (resistance > 18 MΩ/cm) was used throughout the experiments.

Table 1. The oligonucleotides used in this work

DNAs	Sequence[a] (5′→3′)	Extinction coefficient [L/mol·cm]
C-MYC	ACAATTAATACGCTCACTATA**GGGTGGGGA GGGTGGGG**AAGAGCTATGATGCGTTCGAT	593600
C-MYC-c	ATCGAACGCATCATAGCTCTTCCCCACCCTCC CCACCCTATAGTGAGCGTATTAATTGT	547300
NRAS	AATTAATACGACTCACTATA**GGGAGGGCGG GTCTGGG**AAGAGCTATGATGCGTTCGA	573000
NRAS-c	TCGAACGCATCATAGCTCTTCCCAGACCCGCC CCTCCCTATAGTGAGTCGTATTAATT	537500
C-MYC-M1	ACAATTAATACGCTCACTATAGTGTGGTGAGT GTGTGGAAGAGCTATGATGCGTTCGAT	588000
C-MYC-M1-c	ATCGAACGCATCATAGCTCTTCCACACACTCA CCACACTATAGTGAGCGTATTAATTGT	568100
C-MYC-M2	ACAATTAATACGCTCACGTGAGTGTGGAGAG AGGTGGCGTGAGAGATGATGCGTTCGAT	595400
C-MYC-M2-c	ATCGAACGCATCATCTCTCACGCCACCTCTCT CCACACTCACGTGAGCGTATTAATTGT	543700
NRAS-M1	AATTAATACGACTCACTATAGTGAGGTGCGT GTCTGTGAAGAGCTATGATGCGTTCGA	558300
NRAS-M1-c	TCGAACGCATCATAGCTCTTCACAGACACGC ACCTCACTATAGTGAGTCGTATTAATT	220400
NRAS-M2	AATTAATACGACTCACGTGAGAGTGGAGCGC GTGTGAGGAGAGCTATGATGCGTTCGA	584500
NRAS-M2-c	TCGAACGCATCATAGCTCTCCTCACACGCGCT CCACTCTCACGTGAGTCGTATTAATT	533500
Sh-C-MYC	ACTATA**GGGTGGGGAGGGTGGGG**AAGAG	295900
Sh-NRAS	ACTATA**GGGAGGGGCGGGTCTGGG**AAGAG	299000

[a] Sequences in bold and underlined are G-quadruplex-forming core G-rich sequences.

UV-vis absorption spectroscopy and fluorescence spectroscopy

Absorption spectra were measured on a TU-1901 UV-vis spectrophotometer with 1cm-path-length micro quartz cell (40 μL, Starna Brand, England). Fluorescence spectra were measured on a SHIMADZU RF-5301PC spectrofluorimeter with 1cm-path-length micro quartz cell (40 μL, Starna Brand, England). Solutions containing 10 μM individual DNAs, 10 mM Tris-HCl buffer (pH = 7.4), 140 mM KCl, 15 mM NaCl, 1 mM Na_2EDTA were prepared (for the experiments conducted under molecular crowding conditions, 400 mL/L PEG 200 and 0.5 mM $MgCl_2$ were also added). Each solution was heated at 95 °C for 5 min to remove any aggregates, then cooled rapidly to 25 °C and was allowed to incubate at 25 °C for 30 min. After overnight incubation at 4 °C, 5 μM of TMPipEOPP was added, absorption spectra in the range of 350 ~ 800 nm and corresponding fluorescence spectra were recorded. When 454 nm was used as the excitation wavelength, the excitation slit and emission slit were both set at 5 nm. When 700 nm was used as the excitation wavelength, the excitation slit and emission slit were both set at 3 nm (the same below).

UV-vis absorption and fluorescence titration experiments

Titration experiments were carried out by maintaining the G-rich strand concentration constant at 10 μM but varying their individual complimentary sequence concentrations from 0 to 60 μM. The sample solutions were prepared as aforementioned, absorption

spectra in the range of 350 ~ 800 nm and corresponding fluorescence spectra were recorded.

Time-dependent absorption and fluorescence spectra analysis

In the first condition, solutions containing 10 μM individual G-rich sequences, 10 μM complimentary sequences, 10 mM Tris–HCl buffer (pH = 7.4), 140 mM KCl, 15 mM NaCl and 1 mM Na_2EDTA were prepared (for the experiments conducted under molecular crowding conditions, 400 mL/L PEG 200 and 0.5 mM $MgCl_2$ were also added). The solutions were heated at 95 °C for 5 min to remove any aggregates. After cooling rapidly to room temperature, 5 μM of TMPipEOPP was added quickly, and corresponding absorption and fluorescence spectrum signals were recorded at different time points.

In the second condition, 5 μM TMPipEOPP was added in the initial solutions instead of complimentary sequences. After the solutions were cooled rapidly from 95 °C to room temperature, 10 μM of complimentary sequences was added quickly, and corresponding absorption and fluorescence spectrum signals were recorded at different time points.

In the third condition, solutions were prepared as in the first condition. After the solutions were cooled rapidly from 95 °C to room temperature, 5 μM of TMPipEOPP was added. Then, water or PEG 200 was added into the solution, and corresponding absorption and fluorescence spectrum signals were recorded at different time points. To avoid the DNA concentration change before and after PEG 200 addition, two reaction solutions were prepared for each experiment. Water was added instead of PEG 200 in the first one, and it was used to record the initial absorption and fluorescence signals. PEG 200 was added to the second one, this one was used to record PEG 200-triggered time-dependent absorption and fluorescence signal changes. The two reaction solutions have identical final volumes. To avoid evaporation of any water, the solutions were sealed with liquid paraffin in all of above three experiments.

Circular dichroism (CD) spectroscopy

CD spectra were measured on a Jasco J-715 spectropolarimeter. 3 mL reaction mixture was prepared in 10 mM Tris-HCl buffer (pH = 7.4) containing 1 μM individual DNA oligonculeotides, 140 mM KCl, 15 mM NaCl, 1 mM Na_2EDTA were prepared (for the experiments conducted under molecular crowding conditions, 400 mL/L PEG 200 and 0.5 mM $MgCl_2$ were also added). The mixture was heated at 95 °C for 5 min, cooled slowly to 25 °C and then incubated at 4 °C overnight. CD spectrum of the mixture was recorded between 230 and 320 nm in 1 mm path length cuvettes. Spectra were averaged from 3 scans, which were recorded at 100 nm/min with a response time of 1 s and a bandwith of 1.0 nm.

Results and Discussion

UV-vis absorption of TMPipEOPP was used to investigate G-quadruplex/duplex competition embedded in duplex DNAs.

The aim of this work was to develop a simple optical method to investigate the competition between G-quadruplexes and duplexes formed by core G-rich sequences embedded in duplex DNA. The key was to find a G-quadruplex-specific optical probe that could signal the formation of G-quadruplexes *vs.* duplexes or single-stranded DNAs. In our previous work,[24,25] we demonstrated that TMPipEOPP, a cationic porphyrin derivate with four large side arms, was a G-quadruplex-specific optical probe. TMPipEOPP specifically recognizes G-quadruplexes *vs.* duplexes or single-stranded DNAs under both dilute and molecular crowding conditions. To investigate whether or not TMPipEOPP could be used to indicate G-quadruplex formation inside long duplex DNA, two long DNA sequences (NRAS and C-MYC, Table 1) from human genomic DNA, were used. In these two sequences, the potential G-quadruplex-forming core G-rich sequences were in the middle, with flanking sequences that could hybridize with individual complementary sequences to form duplexes. The UV-vis absorption spectrum of free TMPipEOPP showed a strong Soret absorption peak around 425 nm under both dilute and molecular crowding conditions (Fig. 1 and Fig. S1). In the presence of C-MYC or NRAS, the 425 nm absorption signal decreased and new absorption peaks appeared around 454 nm and 694 nm. These results were consistent with those in the presence of G-quadruplexes formed by short G-rich sequences,[24] suggesting that the core G-rich sequences in long DNA molecules folded into G-quadruplex structures.

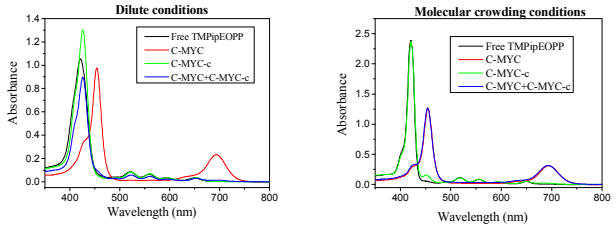


Fig. 1 Effects of C-MYC, C-MYC-c, and their mixture on the absorption spectrum of TMPipEOPP under dilute or molecular crowding conditions**.** The results for NRAS are shown in Fig. S1.

In contrast, when G-rich sequences were replaced by individual complementary sequences (NRAS-c or C-MYC-c), different spectral changes were observed. No new peaks emerged and instead of obvious hypochromicity caused by NRAS or C-MYC, little hypochromicity or even hyperchromicity was observed for the Soret peak at 425 nm in the presence of NRAS-c or C-MYC-c. The complementary strands could not fold into G-quadruplex structures; thus, these results suggested that TMPipEOPP specifically discriminated between G-quadruplexes and single-stranded DNA, and could be used to probe the formation of G-quadruplexes by core G-rich sequences in long DNA molecules. To investigate G-quadruplex/duplex competition in flanking duplex DNA, G-rich sequences were incubated with complementary sequences to form hybrids and the effect on the absorption spectrum of TMPipEOPP was investigated. Under dilute conditions, the absorption peaks at 454 nm and 694 nm seen in the presence of G-rich sequences nearly disappeared with the addition of the complementary sequences and the absorption signal at 425 nm was recovered. These results suggested that core G-rich sequences flanked by duplexes converted to duplex structures. Using poly(ethylene glycol) with an average molecular weight of 200 daltons (PEG 200) as a molecular crowding agent,[26] G-quadruplex/duplex competition was studied under molecular crowding conditions. The presence of complementary sequences had almost no effect on the changes in the TMPipEOPP absorption spectrum induced by G-rich sequences (Fig. 1 and Fig. S1). This result suggested that G-quadruplexes formed even if they were embedded in flanking duplex structures. These results were consistent with previous reports that dilute conditions aid the formation of duplexes and that molecular crowding promotes the formation of G-quadruplexes.[27,28] To further demonstrate this, the effect of PEG 200 concentration on the G-quadruplex/duplex competition was instigated. The results showed that the increase of PEG 200 concentration is benefit for the conversion of duplexes to G-quadruplexes (Fig. S2).

Above results suggested that TMPipEOPP could be used to detect the formation of G-quadruplexes by core G-rich sequences embedded in duplex structures. To further demonstrate the

availability of the method, core G-rich sequences were replaced by other sequences without G-quadruplex-forming potential and the effects on the TMPipEOPP absorption spectrum were investigated (Fig. 2 and Fig. S3). None of the resulted mutants (C-MYC-M1, C-MYC-M2, NRAS-M1 and NRAS-M2, Table 1), their complementary sequences (C-MYC-M1-c, C-MYC-M2-c, NRAS-M1-c and NRAS-M2-c, Table 1), or mixtures of the mutants and individual complementary sequences resulted in the appearance of the two peaks seen when sequences with G-quadruplex-forming potential were used. This result indicated that the emergence of the two peaks at 454 and 694 nm really resulted from the formation of G-quadruplex structures. These results also demonstrated the recognition specificity of TMPipEOPP for G-quadruplexes *vs.* duplex DNAs, supporting the use of this method for studying G-quadruplex/duplex competition in long DNA molecules. Although a weak peak was observed at 454 nm under molecular crowding conditions, the signal intensity was very low and did not affect the judgement of experimental results (Fig. 2 and Fig. S3).

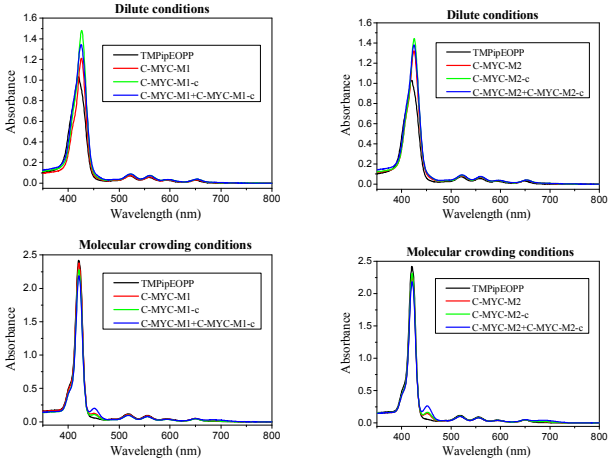


Fig. 2 Effects of the mutants of C-MYC, their complimentary sequences, and mixtures of the mutants and complementary sequences on the absorption spectrum of TMPipEOPP under dilute or molecular crowding conditions. The results for the mutants of NRAS are shown in Fig. S3.

Fluorescence of TMPipEOPP was used to investigate G-quadruplex/duplex competition embedded in duplex DNAs.

Above experiments demonstrated that G-quadruplex structures formed by core G-rich sequences with flanking duplex DNA could be recognized by changes in the UV-vis absorption spectrum of TMPipEOP. Therefore, we investigated whether or not TMPipEOPP could be used as a fluorescent probe for similar assays. Our previous work demonstrated that G-quadruplexes change the fluorescence spectrum of TMPipEOPP but duplexes and single-stranded DNAs do not.[24,25] Free TMPipEOPP had a maximum excitation wavelength at 425 nm, which shifted to 454 nm in the presence of G-quadruplexes. In addition, a new excitation wavelength with a stronger excitation efficiency emerged around 700 nm. When 454 nm was used as the excitation wavelength, free TMPipEOPP emitted weak fluorescence around 660 nm (Fig. 3 and Fig. S4). In dilute conditions, addition of single-stranded DNAs with no G-quadruplex-forming ability, for example C-MYC-c or NRAS-c, slightly increased the fluorescence signal intensity of TMPipEOPP but had almost no effect on the shape of the fluorescence spectrum; maximum emission wavelength was still observed around 660 nm. However, in the presence of NRAS or C-MYC, the maximum emission wavelength shifted to around 720 nm and fluorescence intensities were greatly enhanced. These results were consistent with those in the presence of G-quadruplexes formed by short G-rich sequences,[24] demonstrating that G-quadruplex structures were formed by the two DNA strands. Further addition of complementary sequences (C-MYC-c or NRAS-c) decreased the fluorescence intensity at 720 nm and increased the fluorescence intensity at 660 nm, thus suggesting that G-rich sequences tended to convert from G-quadruplexes to duplexes in the presence of complementary sequences. This is also consistent with the UV-vis absorption spectroscopy result. With excitation at 700 nm, free TMPipEOPP and its mixture with oligonucleotides that could not form G-quadruplexes emitted almost no fluorescence. In contrast, strong fluorescence emission was observed for a mixture of TMPipEOPP and G-rich sequences. Addition of strands complementary to the G-rich sequences led to reduction of fluorescence. The decrease in signal was larger than the decrease seen when 454 nm was the excitation wavelength. Since C-MYC-c and NRAS-c also slightly increased fluorescence at 720 nm when 454 nm was the excitation wavelength, the observation of G-quadruplex/duplex competition may be interfered to some degree. Thus, excitation at 700 nm might be more suitable for probing the formation of G-quadruplexes in duplex DNA.

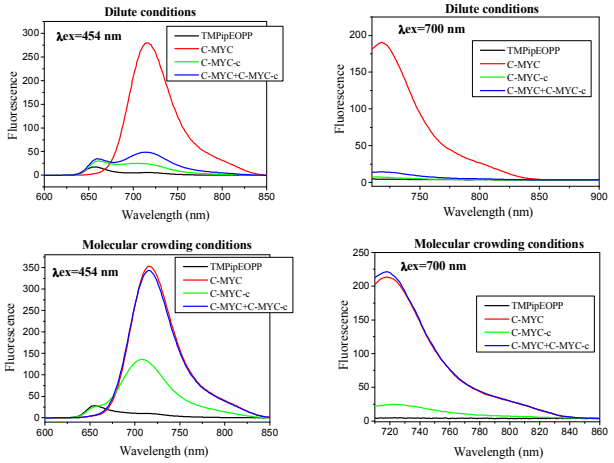


Fig. 3 Effects of C-MYC, C-MYC-c, and their mixture on the fluorescence spectrum of TMPipEOPP under dilute or molecular crowding conditions. The results for NRAS are shown in Fig. S4.

Under conditions of molecular crowding, interference from single-stranded DNA was more pronounced when 454 nm was used as the excitation wavelength. In the presence of 400 mL/L PEG 200, G-rich sequences and mixtures of G-rich sequences and complementary sequences showed similar enhancement of TMPipEOPP fluorescence (Fig. 3 and Fig. S4). These results were consistent with the results from UV-vis spectroscopy. However, because C-MYC-c and NRAS-c also increased fluorescence at 720 nm, we could not conclude that formation of G-quadruplex structures was unaffected by complementary strands. However, when excited at 700 nm, fluorescence emission of TMPipEOPP in the presence of C-MYC-c or NRAS-c was much weaker than fluorescence in the presence of G-rich sequences. This result indicated that when 700 nm was the excitation wavelength, non-G-quadruplex-forming sequences interfered less with the detection of G-quadruplexes in long DNA strands than when 454 nm was the excitation wavelength. The addition of complementary sequences had

little effect on fluorescence of TMPipEOPP/G-rich oligonucleotides mixtures, demonstrating that G-quadruplex structures were still formed in mixtures of G-rich and complementary sequences, consistent with the UV-vis spectroscopy results.

These experiments suggested that TMPipEOPP could be used as a fluorescent probe to indicate the formation of G-quadruplexes by G-rich sequences in duplex DNA. Better results might be obtained using 700 nm as the excitation wavelength. To further explore this possibility, the effects on TMPipEOPP fluorescence of the four C-MYC and NRAS mutants that could not fold into G-quadruplexes were also investigated. In dilute conditions with excitation at 454 nm, the mutants, their complementary sequences, and mixtures of the mutants and complementary sequences almost showed little effect on the shape of the TMPipEOPP fluorescence spectrum (Fig. 4 and Fig. S5). The maximum fluorescence peak still appeared at 660 nm and fluorescence intensity at 720 nm was much lower than intensity in the presence of G-quadruplexes. With excitation at 700 nm, almost no fluorescence was observed in the presence of non-quadruplex-forming single-stranded DNAs or duplexes formed by them. These results further demonstrated that TMPipEOPP efficiently discriminated G-quadruplexes from duplexes and single-stranded DNAs and could be used to probe the formation of G-quadruplexes in long duplex DNAs.

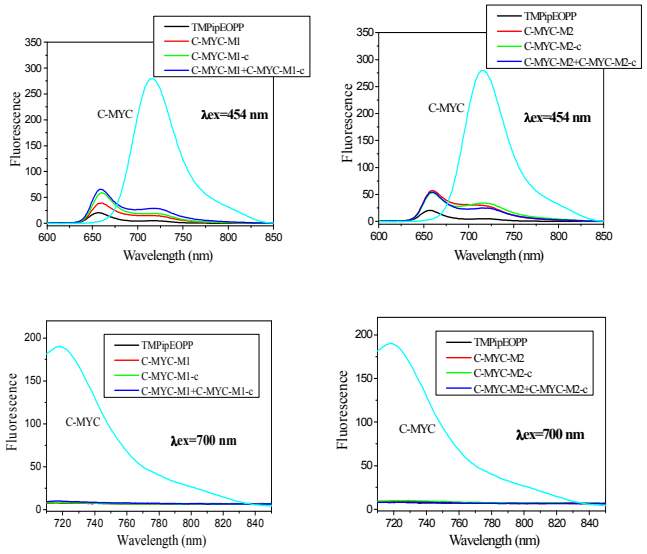


Fig. 4 Effects of the mutants of C-MYC, their complimentary sequences, and mixtures of the mutants and complementary sequences on the fluorescence spectrum of TMPipEOPP under dilute conditions. The results for the mutants of NRAS are shown in Fig. S5.

Under molecular crowding conditions, with excitation at 454 nm, obvious fluorescence signal increase could be observed at 720 nm in the presence of the mutant sequences, their complementary sequences, or mixtures of the mutants and individual complementary sequences (Fig. 5 and Fig. S6). This result was accompanied by changes in the fluorescence spectrum shape, with the maximum wavelength changing to 720 nm. These results suggested that duplexes and single-stranded DNAs might interfere with the detection of G-quadruplexes in duplex DNAs. The interference from duplexes and single-stranded DNAs was reduced when 700 nm was used as the excitation wavelength, thus making TMPipEOPP possible to detect the formation of G-quadruplexes in the presence of duplexes and single-stranded DNAs under conditions of molecular crowding.

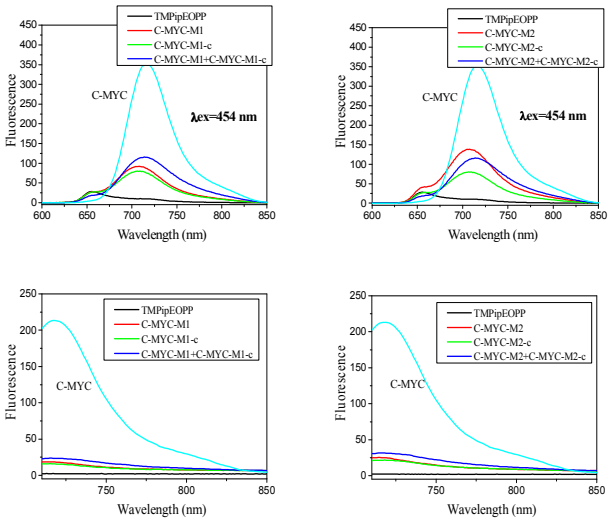


Fig. 5 Effects of the mutants of C-MYC, their complimentary sequences, and mixtures of the mutants and complementary sequences on the fluorescence spectrum of TMPipEOPP under molecular crowding conditions. The results for the mutants of NRAS are shown in Fig. S6.

Complementary strand concentration-dependent changes in absorption and fluorescence of TMPipEOPP/G-rich sequence mixtures.

To investigate whether or not the duplex-forming sequences at two ends of the core G-rich sequences were long enough to embed the G-rich core in duplex structures, DNA concentration-dependent changes in the UV-vis absorption and fluorescence of TMPipEOPP were investigated (Fig. 6 and Fig. S7). In the presence of 10 μM C-MYC or NRAS, TMPipEOPP showed a maximum absorption peak around 454 nm and gave a maximum fluorescence peak around 720 nm when excited at 700 nm. Under dilute conditions with complementary sequences, both absorption and fluorescence decreased. When the concentration ratio of complementary sequences to G-rich sequences was 1:1, absorption and fluorescence were nearly comparable to the results for a 6:1 mixture. It is well known that the stability of intermolecular duplexes increases with DNA concentration.[29] Our results demonstrated that the duplex-forming sequences flanking the G-rich regions were long enough to form stable duplex structures at a complementary/G-rich sequence concentration ratio of 1:1, making the study of G-quadruplex/duplex competition using our proposed method possible. Under molecular crowding conditions, both absorption and fluorescence signals were nearly unchanged as the concentration of complementary sequences increased, even at a ratio of complementary to G-rich sequences of 6:1. These results indicated that G-quadruplex structures formed when flanked by duplex structures under molecular crowding conditions. K^+ concentration nearly had no effect on the competitive behaviour between G-quadruplex and duplex structures under both dilute and molecular crowding conditions (Fig. S8).

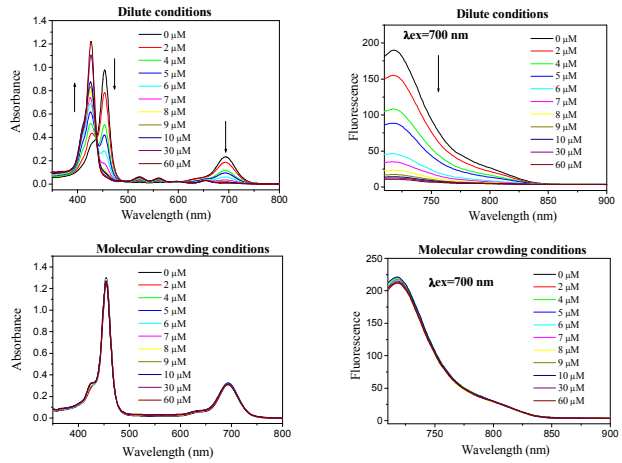


Fig. 6 C-MYC-c concentration-dependent UV-vis absorption and fluorescence spectrum changes of TMPipEOPP/C-MYC mixture under dilute or molecular crowding conditions. The concentrations of C-MYC-c are shown in the figures. The results for NRAS-c are shown in Fig. S7.

Monitoring G-quadruplex/duplex competition in real time.

Compared to gel electrophoresis methods for detecting G-quadruplexe/duplex competition**,** our method is simple, easy and could be used to continuously monitor G-quadruplex/duplex competition in real time. Using our method, G-quadruplex/duplex competition was monitored in real time under three conditions. In the first, G-rich sequences with single-stranded structures mixed with complementary sequences. *In vivo*, DNA duplexes locally and transiently dissociate to form single-stranded DNA during events such as replication, transcription and promoter recognition.[11] To generate single-stranded DNAs, mixtures of G-rich sequences and complementary sequences were heated at 95 ^{o}C for 5 min to destroy any folded structures. Then, the mixtures were cooled to room temperature rapidly to trigger the competition between G-quadruplex and duplex structures.[11] After addition of TMPipEOPP, absorption and fluorescence changes were monitored. Under dilute conditions, both initial absorption (at 454 nm) and initial fluorescence were very low (Fig. S9). Over time, signals were nearly unchanged, indicating that intermolecular duplexes formed rapidly, were stable and did not disassociate with time. In contrast, under molecular crowding conditions, high levels of initial absorption (454 nm) and fluorescence were observed (Fig. S10) and were unchanged with time. These results suggested that G-quadruplex structures formed and were stable over time.

In the second condition, G-rich sequences were allowed to fold into G-quadruplex structures, then complementary sequences were added to trigger competition between G-quadruplexes and duplexes. G-rich sequences were heated and after cooling to room temperature, then, TMPipEOPP and complementary sequences were added and absorption and fluorescence were monitored. High initial absorption (454 nm) and fluorescence signals were observed under both dilute and molecular crowding conditions, suggesting that G-quadruplex structures formed under both conditions. Under dilute conditions, with the addition of complementary sequences, absorption and fluorescence decreased with time (Fig. 7 and Fig. S11). However, even after 7.5 days, obvious absorption and fluorescence were still observed, indicating that some G-quadruplexes remained in the mixtures. These results, combining with those obtained in the first condition, indicated that unstructured G-rich sequences rapidly hybridized with complementary sequences to form stable duplex structures under dilute conditions. However, when G-rich sequences fold into G-quadruplexes, addition of complementary sequences promoted conversion from G-quadruplexes to duplexes at a slow conversion ratewith some G-rich sequences remaining as G-quadruplex structures. One reason may be that the conversion needs the disruption of G-quadruplex structures, and the stability of G-quadurplex structures increases the difficulty of structural conversion. In contrast, under molecular crowding conditions, no time-dependent absorption or fluorescence changes were observed with the addition of complementary sequences (Fig. 7 and Fig. S11). This result was consistent with previous reports that G-quadruplexes are more stable under molecular crowding conditions.[18-20,30] Because of the stability of G-quadruplexes, they did not convert to duplexes in the presence of complementary sequences. To assess the potential effects of TMPipEOPP on the competition between G-quadruplex and duplex, we designed another experiments. That is, TMPipEOPP was not included in the mixture of preformed G-quadruplexes and complementary sequences, but was added at different time points. After TMPipEOPP addition, the absorption and fluorescence spectra of the mixtures were recorded immediately. Similar spectral changes were obtained for the mixtures, in which TMPipEOPP was pre-added or not (Fig. S12), indicating the presence of TMPipEOPP has little effects on the G-quadruplex/duplex competition.

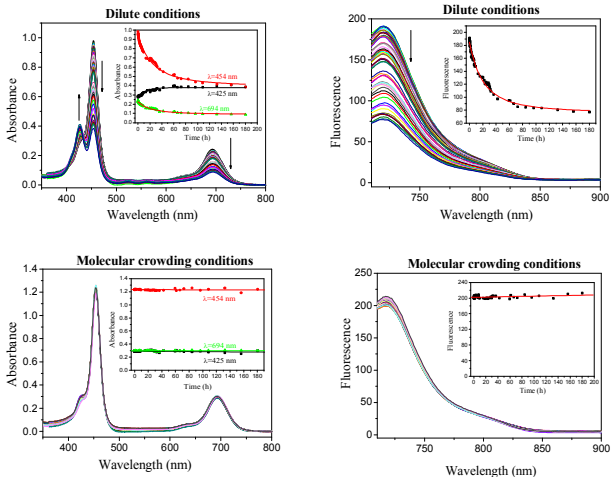


Fig. 7 Time-dependent G-quadruplex/duplex competition in flanking duplex DNAs under dilute or molecular crowding conditions. In this experiment, C-MYC was allowed to form G-quadruplex structures before addition of C-MYC-c. The inserts showed time-dependent absorption signal changes at 425, 454, 694 nm, and fluorescence signal changes at 720 nm. Corresponding results for NRAS are shown in Fig. S11.

Since different G-quadruplex/duplex competition results were observed under dilute and molecular crowding conditions, we investigated the effect of PEG 200 on competition as the third condition (Fig. 8 and Fig. S12). G-rich sequences tend to form duplexes in the presence of complementary sequences under dilute conditions, which was reflected by strong absorption at 425 nm, low absorption at 454 nm and 694 nm, and low fluorescence at 720 nm. Addition of PEG 200 rapidly decreased absorption at 425 nm with an increase in absorption at 454 and 694 nm and fluorescence at 720 nm. Then, absorption at 425 nm decreased slightly and absorption at 454

nm and 694 nm and fluorescence at 720 nm increased slightly with time. These results suggested that PEG 200 promoted rapid structural conversion of G-rich sequences from duplexes to G-quadruplexes.

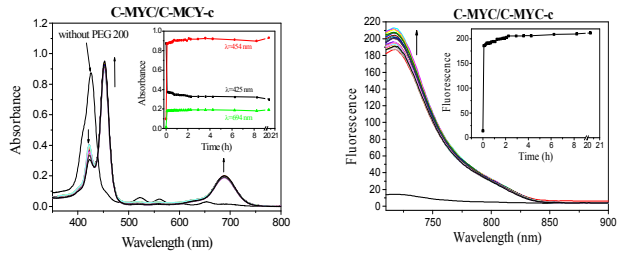


Fig. 8 Time-dependent G-quadruplex/duplex competition in C-MYC/C-MYC-c mixture after PEG 200 addition. In this experiment, duplex structures were formed under dilute conditions, then, PEG 200 was added to trigger G-quadruplex/duplex competition. The inserts showed time-dependent absorption signal changes at 425, 454, 694 nm, and fluorescence signal changes at 720 nm. The results for NRAS/NRAS-c mixture are shown in Fig. S13.

Comparison to circular dichroism spectroscopy for G-quadruplex/duplex competition.

Circular dichroism (CD) spectroscopy is commonly used to probe the formation of G-quadruplexes and study DNA structural changes. A typical parallel G-quadruplex shows a positive CD peak around 260 nm and a negative peak around 240 nm. A typical antiparallel G-quadruplex shows a positive CD peak around 290 nm.[31] Double-stranded or single-stranded DNA shows a positive peak in the CD spectrum. The wavelength of peak shifts between 260 and 280 nm depending on the sequence.[32,33] CD spectroscopy has been used to study competition between G-quadruplexes and duplexes formed by short G-rich sequences in the presence of complementary short C-rich sequences,[17,18,34] although this technique does not discriminate well between the two structures since the duplex and quadruplex CD spectra partially overlap.[35] As for the studies of the G-rich sequences embedded in long DNA molecules, CD spectroscopy even gives worse results. Using G-quadruplex/duplex competition studies in C-MYC/C-MYC-c mixture as an example (Fig. 9). Under dilute conditions, the CD spectrum of C-MYC-c gave a positive peak around 276 nm, indicating that the oligonucleotide was an unstructured single-stranded coil. The CD spectrum of C-MYC gave a positive peak around 269 nm, suggesting that parallel G-quadruplex might have been formed. However, because of the presence of single-stranded tails at two ends of core G-rich sequence, the wavelength shifted from 263 nm of short core G-rich sequence (Sh-C-MYC, Table 1) to 269 nm. Thus, the CD spectral characteristics of the G-quadruplexes formed in long DNA were weakened, making the distinction between G-quadruplexes and single-stranded or duplex DNAs more difficult. When C-MYC and C-MYC-c were mixed, a positive peak with higher CD intensity was observed around 275 nm. Similar results were obtained under molecular crowding conditions. The CD spectrum of the mixture gave a positive peak with a wavelength between the peaks of free C-MYC and free C-MYC-c, with a CD signal intensity that was higher than free C-MYC and free C-MYC-c. The results of UV-vis absorption and fluorescence spectroscopy demonstrated that C-MYC mainly hybridized with C-MYC-c to form duplex structures under dilute conditions and folded into G-quadruplex structures under molecular crowding conditions even in the presence of C-MYC-c. However, this difference was not reflected in the CD spectrum change, indicating that CD spectroscopy was not suitable for investigating the competition between G-quadruplexes and duplexes embedded in long duplex DNAs. The optical probe TMPipEOPP might be a better choice for this type of assay. A similar conclusion was drawn from results with other G-rich oligonucleotide NRAS (Fig. S14).

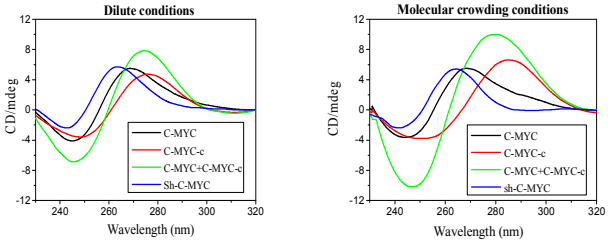


Fig. 9 CD spectra of C-MYC, C-MYC-c and their mixture under dilute or molecular crowding conditions. Sh-C-MYC is used as a control. The results for NRAS are shown in Fig. S14.

Conclusions

In summary, a simple optical method was developed to probe the formation of G-quadruplex structures inside long duplex DNA. This method used the optical probe TMPipEOPP, which specifically recognized G-quadruplexes vs. duplexes and single-stranded DNAs under both dilute and molecular crowding conditions. By following changes in UV-vis absorption and fluorescence spectra, competition between G-quadruplexes and duplex structures was monitored. Results were dependent on conditions. Under dilute conditions, the stability of the duplex structure was higher than the G-quadruplex structure and G-quadruplexes were not formed by core G-rich sequences within duplex structures. Under molecular crowding conditions, however, the stability of G-quadruplex structures increased and their formation was observed. Using TMPipEOPP as an optical probe allowed continuously monitoring of G-quadruplex/duplex competition. Under dilute conditions, unstructured G-rich sequences rapidly hybridized with complementary sequences to form stable duplexes that did not disassociate with time. If complementary sequences were mixed with pre-folded G-quadruplexes, the G-quadruplexes slowly converted to duplex structures. Under molecular crowding conditions, stable G-quadruplexes formed regardless of the presence of complementary sequences and did not disassociate with time. Changing solution conditions from dilute to molecular crowding promoted structural conversion of G-rich sequences from duplexes to G-quadruplexes. This work could provide an important tool for studying G-quadruplexes, which could provide important information for the formation of G-quadruplexes *in vivo*.

Acknowledgements

This work was supported by the Natural Science Foundation of China (No. 21371130) and the National Basic Research Program of China (No. 2011CB707703).

Notes and references

Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), and Department of Chemistry, Tianjin University, Tianjin, 300072, P R China. Fax: (+)86-22-27403475. E-mail: linazhu@tju.edu.cn

Electronic Supplementary Information (ESI) available: [details of corresponding experiments for NRAS and its mutants, effects of PEG 200 or K^+ concentration on the G-quadruplex/duplex competition, monitoring G-quadruplex/duplex competition in real time]. See DOI: 10.1039/b000000x/

1 N. Saini, Y. Zhang, K. Usdin and K. S. Lobachev, *Biochimie*, 2008, **95**, 117.
2 M. Tarsounas and M. Tijsterman, *J. Mol. Biol.*, 2013, **425**, 4782.
3 O. Doluca, J. M. Withers and V. V. Filichev, *Chem. Rev.*, 2013, **113**, 3044.
4 M. L. Bochman, K. Paeschke and V. A. Zakian, *Nat. Rev. Genet.*, 2012, **13**, 770.
5 L. Petraccone, *Top. Curr. Chem.*, 2013, **330**, 23.
6 J. Eddy and N. Maizels, *Nucleic Acids Res.*, 2008, **36**, 1321.
7 M. Düchler, *J. Drug Target.*, 2012, **20**, 389.
8 D. Yang and K. Okamoto, Future Med. Chem., 2010, 2, 619.
9 D. Castillo-González, G. Pérez-Machado, F. Pallardó, T. M. Garrigues-Pelufo and M. A. Cabrera-Pérez, *Curr. Top. Med. Chem.*, 2012, **12**, 284.
10 S. Nakayama, I. Kelsey, J. Wang and H. O. Sintim, *Chem. Commun.*, 2011, **47**, 4766-4768.
11 K. W. Zheng, Z. Chen, Y. H. Hao and Z. Tan, *Nucleic Acids Res.*, 2010, **38**, 327.
12 W. Li, P. Wu, T. Ohmichi and N. Sugimoto, *FEBS Lett.*, 2002, **526**, 77.
13 W. Li, D. Miyoshi, S. Nakano and N. Sugimoto, *Biochemistry*, 2003, **42**, 11736.
14 L. Lacroix, A. Séosse and J. L. Mergny, *Nucleic Acids Res.*, 2011, **39**, e21.
15 J. L. Zhang, Y. Fu, L. Zheng, W. Li, H. Li, Q. Sun, Y. Xiao and F. Geng, *Nucleic Acids Res.*, 2009, **37**, 2471.
16 Z. Y. Kan, Y. Yao, P. Wang, X. H. Li, Y. H. Hao and Z. Tan, *Angew. Chem. Int. Ed. Engl.*, 2006, **45**, 1629.
17 Z. Y. Kan, Y. Lin, F. Wang, X. Y. Zhuang, Y. Zhao, D. W. Pang, Y. H. Hao and Z. Tan, *Nucleic Acids Res.*, 2007, **35**, 3646.
18 D. Miyoshi, S. Matsumura, S. Nakano and N. Sugimoto, *J. Am. Chem. Soc.*, 2004, **126**, 165.
19 D. Miyoshi, H. Karimata and N. Sugimoto, *Angew. Chem. Int. Ed. Engl.*, 2005, **29**, 3740.
20 A. Arora and S. Maiti, *J. Phys. Chem. B* 2009, **113**, 8784.
21 Z. Yu, J. D. Schonhoft, S. Dhakal, R. Bajracharya, R. Hegde, S. Basu and H. Mao, *J. Am. Chem. Soc.*, 2009, **131**, 1876.
22 J. Jaumot, R. Eritja, T. Tauler and R. Gargallo, *Nucleic Acids Res.*, 2006, **34**, 206.
23 N. Kumar and S. Maiti, *Nucleic Acids Res.*, 2005, **33**, 6723.
24 L.-N. Zhu, S.-J. Zhao, B. Wu, X.-Z. Li and D.-M. Kong, *PLoS ONE* 2012, **7**, e35586.
25 L.-N. Zhu, B. Wu and D.-M. Kong, *Nucleic Acids Res.*, 2013, **41**, 4324.
26 D. Miyoshi, T. Fujimoto and N. Sugimoto, *Top. Curr. Chem.*, 2013, **330,** 87.
27 D. Miyoshi, H. Karimata and N. Sugimoto, *J. Am. Chem. Soc.*, 2006, **128**, 7957.
28 S. Nakano, H. Karimata, T. Ohmichi, J. Kawakami and N. Sugimoto, *J. Am. Chem. Soc.*, 2004, **126**, 14330.
29 P. Alberti and J.-L. Mergny, *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 1569.
30 S. Dhakal, Y. Cui, D. Koirala, C. Ghimire, S. Kushwaha, Z. Yu, P. M. Yangyuoru and H. Mao, *Nucleic Acids Res.*, 2013, **41**, 3915.
31 A. Randazzo, G. P. Spada and M. W. da Sulva, *Top. Curr. Chem.*, 2013, **330**, 67.
32 J. Kypr, I. Kejnovská, D. Renčiuk and M. Vorlíčková, *Nucleic Acids Res.*, 2009, **37**, 1713.
33 S.-M. Jia, X.-F. Liu, P. Li, D.-M. Kong and H.-X. Shen, *Biosens. Bioelectron.* 2011, **27**, 148.
34 J. Zhou, C. Wei, G. Jia, X. Wang, Z. Feng and C. Li, *Biochimie*, 2009, **91**, 1104.
35 L. Petraccone, B. Pagano and C. Giancola, *Methods*, 2012, **57**, 76.